Enhancing Large Urban Photo Collections with 3D Ladar and GIS Data

Peter Cho[1], Noah Snavely[2]

[1]Active Optical Systems Group, MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420, USA

[2]Department of Computer Science, Cornell University, Ithaca, NY 14853 , USA

[1]cho@ll.mit.edu; [2]snavely@cs.cornell.edu

Abstract

Recent work in computer vision has demonstrated the potential to automatically recover camera and scene geometry from large collections of uncooperatively-collected photos. At the same time, aerial ladar and Geographic Information System (GIS) data are becoming more readily accessible. In this paper, we present a system for fusing these data sources in order to transfer 3D and GIS information into outdoor urban imagery.

Applying this system to 1000+ pictures shot of the lower Manhattan skyline and the Statue of Liberty, four proof-of-concept examples of geometry-based photo enhancement are presented which are difficult to perform via conventional image processing: feature annotation, image-based querying, photo segmentation and city image retrieval. In each example, high-level knowledge projects from 3D world-space into georegistered 2D image planes and/or propagates between different photos. Such automatic capabilities lay the groundwork for future real-time labeling of imagery shot in complex city environments by mobile smart phones.

Keywords

Photo Reconstruction; Ladar Map; Georegistration; Knowledge Propagation

Introduction

The quantity and quality of urban digital imagery are rapidly increasing over time. Millions of photos shot by inexpensive electronic cameras in cities can now be accessed via photo sharing sites such as Flickr and Google Images. However, imagery on these websites is generally unstructured and unorganized. It is consequently difficult to relate Internet photos to one another as well as to other information within city scenes. Some organizing principle is needed to enable intuitive navigating and efficient mining of vast urban imagery archives.

Fortunately, 3D geometry provides such an organizing principle for images taken at different times, places and resolutions. Recent work in computer vision has demonstrated automatic capability to recover relative geometric information for uncooperatively-collected images [1]. Moreover, several other urban data sources including satellite views, aerial ladar point clouds and GIS layers can serve as absolute geometrical underlays for such reconstructed photo collections. High-level knowledge such as building names and city object geocoordinates can then be transferred from the underlays into the ground-level imagery via geometrical projection.

In this paper, we demonstrate the utility of 3D geometry for enhancing 2D urban imagery following the flow diagram in Figure 1. Working with uncooperatively collected photos shot in New York City (NYC), we first reconstruct their cameras' relative positions and orientations as well as sparse urban scene geometry. We next fuse aerial ladar, satellite imagery and GIS data to produce a dense three-dimensional NYC map. The reconstructed photos are then georegistered with the 3D map. We illustrate good qualitative alignment between the independent ground-level and aerial data sets and estimate a quantitative registration error.

Once a 3D framework for analyzing 2D photos is established, many challenging urban image enhancement problems become tractable. We focus in particular on automatic feature annotation, image-based querying, photo segmentation and city image retrieval. Examples of geometry-mediated labeling of buildings, measuring of ranges to target points, classifying of sky plus sea regions in photos, and ranking of NYC imagery based upon text search queries are presented. Finally, the study is finished by summarizing our results and discussing future applications of this work.
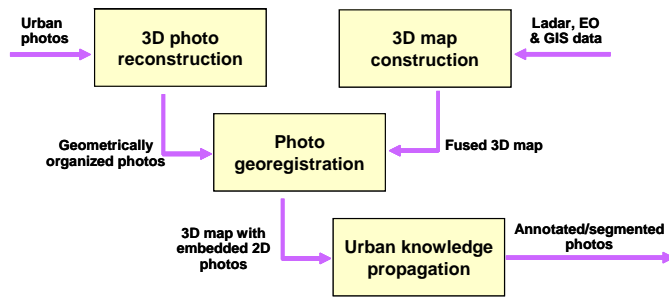
FIG. 1 ALGORITHM FLOW DIAGRAM

Related Work

3D reconstruction and enhancement of urban imagery have been active areas of research over the past decade. We briefly review here some recent, representative articles in this field which overlap our focus.

3D Urban Modeling

Automatic and semi-automatic image-based modeling of city scenes have been investigated by many authors. For example, Xiao et al [2] developed a semi-automatic approach to facade modeling using street-level imagery. Cornelis et al [3] presented a city-modeling framework based upon survey vehicle stereo imagery. And Zebedin et al [4] automatically generated textured building models via photogrammetry methods. While modeling per se is not the focus of our work, we also utilize simple 3D models derived from ladar and structure from motion (SfM) techniques.

Multi-Sensor Urban Data Fusion

Imagery captured from digital cameras has been fused with other data sources in order to construct 3D urban maps in the past. Robertson and Cipolla [5] developed an interactive system for creating models of uncalibrated architectural scenes using map constraints. Pollefeys et al [6] combined video streams, GPS and inertial measurements to reconstruct georegistered building models. Grabler et al [7] merged ground-level imagery with aerial information plus semantic features in order to generate 3D tourist maps.

As ladar technology has matured, many other researchers have incorporated 3D range information into city modeling. For example, Frueh and Zakhor [8] automatically generated textured 3D city models by combining aerial and ground ladar data with 2D imagery. Hu et al [9] developed a semi-automatic modeling system that fuses ladar data, aerial imagery

and ground photos. Zhao et al [10] presented a point cloud registration method which aligns oblique video with 3D sensor data. Ding et al [11] presented an algorithm for registering oblique aerial imagery onto 3D models derived from ladar. Our work also fuses ladar, satellite images and GIS data in order to form a 3D urban map. However, our main contribution is using such rich data to enhance large photo collections.

Urban Digital Photo Georegistration

As digital camera technology has proliferated, interest has grown in georegistering digital photos shot in cities with 3D urban maps. Cho [12] aligned a few photos with ladar data and demonstrated 2D feature identification. Kopf et al [13] developed an interactive system for manually registering individual photos with urban models. Kaminsky et al [14] demonstrated automatic alignment of reconstructed urban photos to satellite imagery. In this paper, we present a semi-automatic approach to align large reconstructed photo sets with ladar data.

Urban Knowledge Propagation

Many different schemes for annotating, querying and retrieving digital urban images have been explored in recent years, with varying degrees of automation. Snavely et al [1] demonstrated automatic transfer of labels from one photo to another. In contrast, Russell et al [15] presented a web-based annotation tool that can be used to manually label objects in large numbers of images. Cho et al [16] demonstrated depth retrieval for video frames aligned with panoramas georegistered with ladar data. Content-based image retrieval has been extensively reviewed by Datta et al [17].

In [13], Kopf et al highlighted a number of applications for information transfer into single photos including image dehazing, relighting and annotation. We go further by demonstrating knowledge propagation into entire photo collections.

The primary contribution of our paper is demonstrating the utility of a geometric framework combining ladar data, GIS layers, and structure from motion for automatic urban imagery enhancement. Outdoor city photo annotation, querying and retrieval all become much more tractable when information anchored to 3D geometry is leveraged.

3D Photo Reconstruction

We begin by downloading more than 1000 digital

photos shot of the lower Manhattan skyline and the Statue of Liberty from www.flickr.com. Flickr's website contains vast numbers of such pictures which users have tagged as generally related to New York City. But our uncooperatively-collected data set is otherwise unorganized. We therefore recover 3D structure from these photo collections using the SfM approach of Snavely *et al.* [1].

In brief, the 3D reconstruction procedure first extracts salient local features from each input image using Scale Invariant Feature Transform (SIFT) features which are designed to be invariant to common geometric and photometric variations [18]. SIFT features can be robustly matched between all pairs of images via nearest-neighbor searching [18-20] plus RANSAC filtering [21].

SIFT feature matching itself begins to impose structure upon an unorganized set of photos. In particular, feature matches define an image connectivity graph [22]. Each photo corresponds to a node in the graph, and edges link images with matching SIFT features. The image graph for 1012 Manhattan photos is shown in Figure 2, using a graph layout and interactive rendering tool developed by M. Yee [23]. For this particular set of 1000+ photos, the graph is divided roughly into two large clusters, representing the Manhattan skyline and the Statue of Liberty.

Yee's viewer enables one to intuitively inspect the graph's substructure. For example, if we zoom into the lower-left cluster of FIG. 2, the circular nodes are replaced by thumbnails of Statue of Liberty photos (see fig. 3a). Similarly, skyline photos are grouped together within the upper-right cluster. Nodes located in the graph's neck region between the two large clusters exhibit both Statue and skyline content as one would expect (see fig. 3b).

Once corresponding features have been extracted and matched between multiple photos, our system next employs structure from motion techniques to recover camera poses and sparse scene structure. SfM takes as input the 2D feature matches and computes a set of 3D scene points, as well as the rotation, position and focal length parameters for each photo. We use the Bundler toolkit to solve the underlying optimization problem [24]. SfM results for the 1012-image collection of NYC Flickr photos are illustrated in FIG. 4.

Given its high computational complexity, 3D reconstruction for large numbers of photos must currently be performed on a parallel cluster. We ran
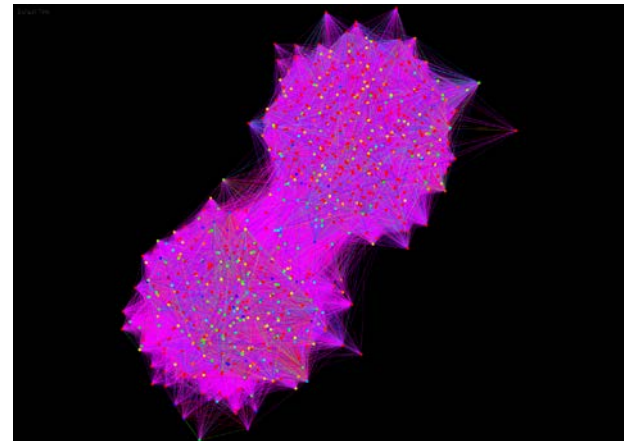

FIG. 2 GRAPH ILLUSTRATING TOPOLOGY FOR 1012 NYC PHOTOS. NODES CORRESPOND TO PHOTOS AND ARE COLORED ACCORDING TO THE ESTIMATED UNCERTAINTY IN THE CORRESPONDING CAMERA LOCATION. SIFT FEATURE OVERLAP BETWEEN PHOTOS IS INDICATED BY THE GRAPH'S COLORED EDGES
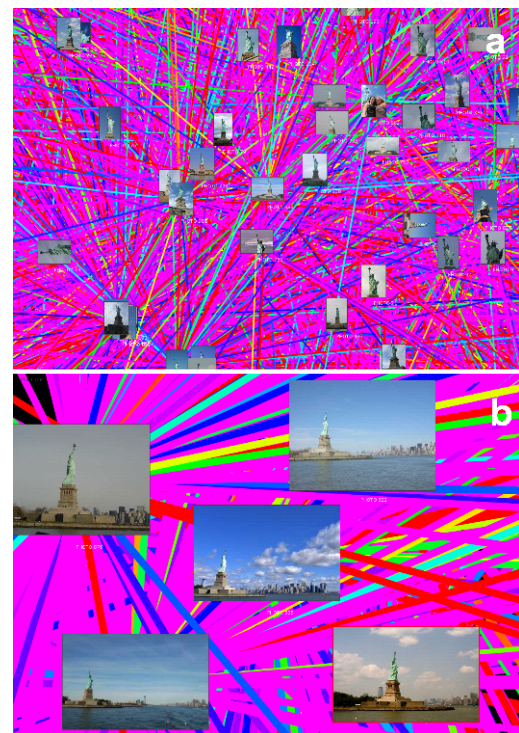

FIG. 3 (A) CIRCULAR NODES IN THE LOWER-LEFT SUBCLUSTER OF FIGURE 2 TURN INTO STATUE OF LIBERTY PHOTO THUMBNAILS WHEN THE USER ZOOMS INTO THE GRAPH. (B) THUMBNAILS IN THE GRAPH'S MIDDLE NECK REGION EXHIBIT BOTH STATUE AND SKYLINE CONTENT
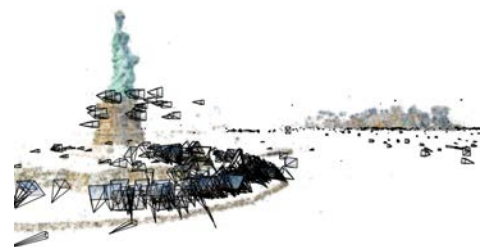

FIG. 4 RELATIVE POSITIONS AND POSES FOR 1012 CAMERAS AUTOMATICALLY DETERMINED BY STRUCTURE FROM MOTION. RELATIVE STATUE AND SKYLINE GEOMETRY ARE ALSO RETURNED WITHIN A SPARSE POINT CLOUD OUTPUT

parallelized versions of the feature extraction and matching steps on Lincoln Laboratory's high-performance, 130-processor Grid [25]. The reconstruction process took approximately 4 hours on this cluster.

Conventional digital cameras only capture angle-angle projections of the 3D world onto 2D image planes. In the absence of metadata, photos yield neither absolute lengths nor absolute distances. It is therefore difficult to automatically determine absolute position, orientation or scale from a set of Internet images. In order to georegister reconstructed photos, we need to incorporate additional sensor data. We therefore construct 3D urban maps based upon aerial ladar data.

3D Map Construction

High-resolution ladar imagery of entire cities is now routinely gathered by platforms operated by government laboratories as well as commercial companies. Airborne laser radars collect hundreds of millions of city points whose geolocations are efficiently stored in and retrieved from multiresolution quadtrees. Ladars consequently yield detailed geospatial underlays onto which other sensor measurements can be draped.

We work with a Rapid Terrain Visualization map collected over New York City on Oct 15, 2001. These data have a 1 meter ground sampling distance. By comparing absolute geolocations for landmarks in this 3D map with their counterparts in other geospatial databases, we estimate these ladar data have a maximum local georegistration error of 2 meters.

Complex urban environments are only partially characterized by their geometry. They also exhibit a rich pattern of intensities, reflectivities and colors. So the next step in generating an urban map is to fuse an overhead image with the ladar point cloud. For our New York example, we obtained Quickbird satellite imagery which covers the same area as the 3D data. Its 0.8 meter ground sampling distance is also comparable to that of the ladar imagery.

We next introduce GIS layers into the urban map. Such layers are commonplace in standard mapping programs which run on the web or as standalone applications. They include points (e.g. landmarks), curves (e.g. transportation routes) and regions (e.g. political zones). GIS databases generally store longitude and latitude coordinates for these geometrical structures, but most do not contain

altitude information. Fortunately, height values can be extracted from the ladar underlay once lateral GIS geocoordinates are known.

After fusing together the ladar map, satellite image and GIS data, we derive the 3D map of New York City presented in FIG. 5. In this map, the hue of each point is proportional to its estimated altitude, while saturation and intensity color coordinates are derived from the satellite imagery. GIS annotations supply useful context.
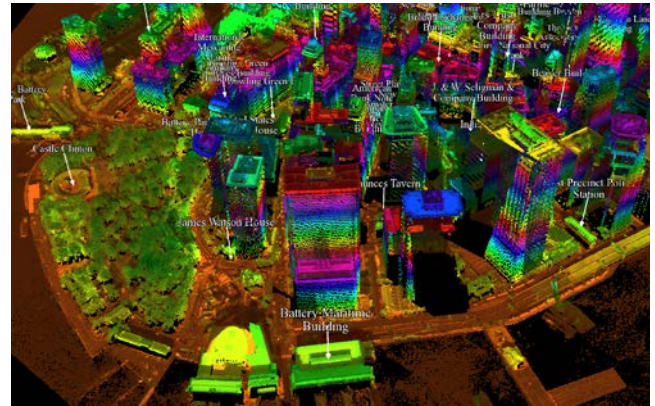

FIG. 5 FUSED 3D MAP OF NEW YORK CITY

Three-dimensional urban maps serve as global backdrops into which information localized in space and/or time may be incorporated. We therefore proceed to combine relative photo reconstructions with absolute coordinates from the NYC map to georegister large numbers of photos.

Reconstructed Photo Georegistration

In order to georegister the SfM reconstruction (in its own relative coordinate system) with the absolute 3D urban map, we select 10 photos with large angular coverage and small reconstruction uncertainties. We then manually pick 33 features in the ladar map coinciding primarily with building corners and identify 2D counterparts to these features within the 10 photos. A least-squares fitting procedure subsequently determines the global transformation parameters needed to align all 1012 reconstructed photos with the ladar map.

This manual step could potentially be automated given GPS information for a subset of photos [14]. As future work, it would also be interesting to rerun the SfM optimization with constraints derived from correspondences between the SfM model and the ladar point cloud. This would help correct for any non-rigid deformations between the two data sources.

FIG. 6 illustrates the reconstructed photos

georegistered with the NYC map. In order to efficiently display large numbers of pictures in our 3D viewer, they are rendered as low-resolution thumbnails inside view frusta when the virtual camera is located far away in space. When the user double clicks on some view frustum, the virtual camera zooms in to look at the full-resolution version of the selected image. For example, FIG. 7 illustrates a Statue of Liberty photo in front of the reconstructed Statue point cloud (for which we do not have ladar data). By comparing geocoordinates for reconstructed points on the Statue of Liberty with their pixel counterparts in Google Earth satellite imagery, we estimate that the average angular orientation error for our georegistered cameras is approximately 0.1 degree.
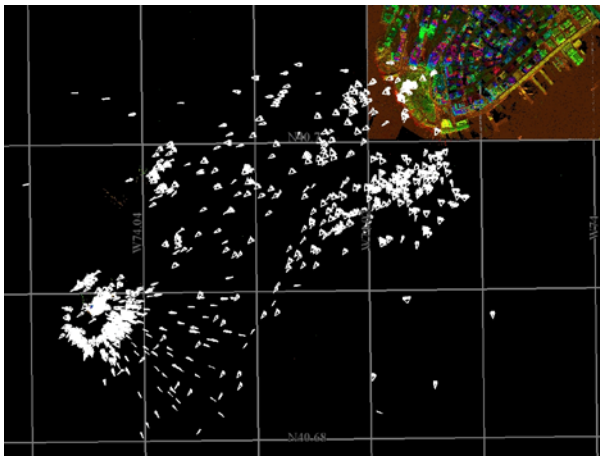

FIG. 6 1012 RECONSTRUCTED PHOTOS GEOREGISTERED WITH THE 3D NYC MAP

A more stringent test of the georegistration accuracy is provided by the alignment between projected ladar points and their corresponding image pixels, particularly for cameras located far away from their subject (e.g., the images of the skyline within FIG. 8). FIG. 9 exhibits the match between one skyline photo and the ladar background. Their agreement represents a nontrivial georegistration between two completely independent data sets. Similar qualitatively good alignment holds for nearly all of the skyline photos and the 3D map.

Urban Knowledge Propagation

Once the reconstructed photo collection is georegistered with the 3D urban map, many difficult image segmentation and enhancement problems become fairly straightforward. In this section, we present four proof-of-concept examples of geometry-based augmentation of geospatially organized photos which would be very difficult to perform via conventional computer vision algorithms.
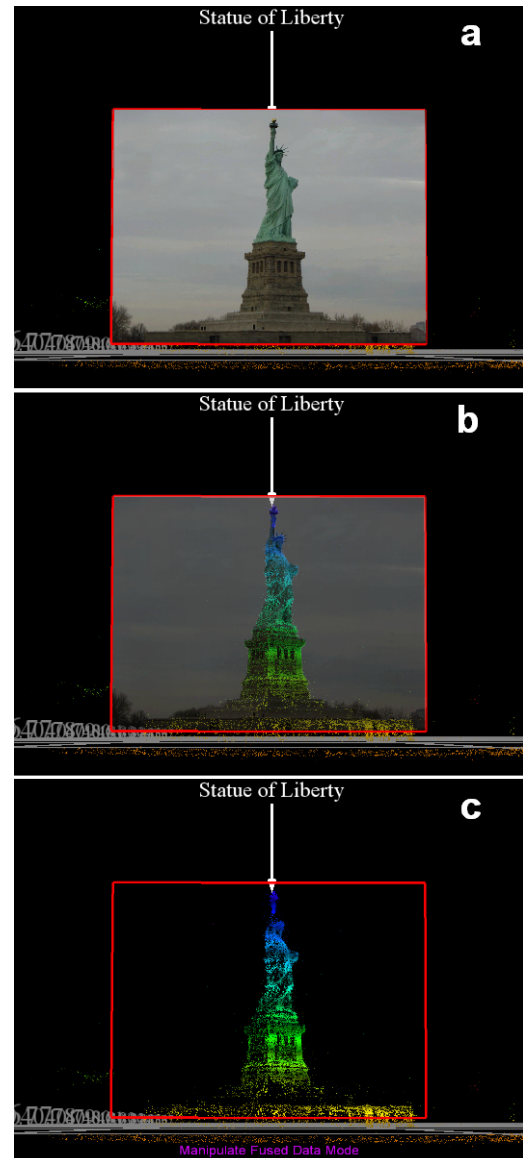

FIG. 7 ONE STATUE PHOTO DISPLAYED IN FRONT OF THE RECONSTRUCTED POINT CLOUD WITH (A) 0%, (B) 50% AND (C) 100% ALPHA BLENDING

FIG. 8 APPROXIMATELY 500 RECONSTRUCTED SKYLINE PHOTOS GEOREGISTERED WITH THE 3D NYC MAP.

Urban Feature Annotation

Our first enhancement application is automatically annotating static features in complex urban scenes. For example, we would like a machine to label buildings

in NYC skyline photos. This annotation problem is extremely challenging from a conventional 2D standpoint due to the wide range of possible viewing and illumination conditions. But once a photo collection is georegistered, we leverage the fact that building names are tied to specific geolocations. After a camera has been globally reconstructed, projecting skyscraper geolocations into its image plane is simple. Skyscraper labels thus transfer automatically from 3D down to 2D. This basic projection approach holds for other information which is geospatially anchored, including roadway networks and political zones.

One technical problem for urban knowledge projection arises from line-of-sight occlusion. To overcome this issue, we convert our ladar point cloud into a height map and assume walls drop straight downwards from rooftop ladar data. If a ray traced from a 3D point back to a reconstructed camera encounters a wall, the point is deemed to be occluded from the camera's view, and information associated with that point is not used to annotate the image. We note that this approach works only for static occluders like buildings and not for transient occluders such as people and cars. Representative building name annotation results from this projection and raytracing procedure are shown in FIG. 10.

Information Transfer Between Images

Our second example demonstrates knowledge propagation between image planes mediated by geometry. FIG. 11 illustrates a prototype image-based querying tool which exhibits the georegistered NYC photos in one window and a particular camera's view in a second window. When a user selects a pixel in the photo on the left, a corresponding 3D point is identified via raytracing in the map on the right. A set of 3D crosshairs marks the world-space counterpart. The geocoordinates and range for the raytraced world-space point are returned and displayed alongside the picked pixel in the 2D window. Note that ocean pixels selected in Figure 11 are reported to lie at 0 meter altitude above sea level, as expected.

Once a 3D point corresponding to a selected 2D pixel is identified, it can be projected into any other camera so long as raytracing tests for occlusion are performed. For instance, the distances from a new camera to previously selected urban features of interest are reported in FIG. 12. Similarly, static counterparts in overlapping air and ground views could be automatically matched. Even handoff of dynamic

urban movers between multiple cameras should be possible provided the movers' length scales are *a priori* known.

Photo Segmentation

Image segmentation represents another application that can be dramatically simplified using a 3D map. For instance, suppose we want to classify every pixel
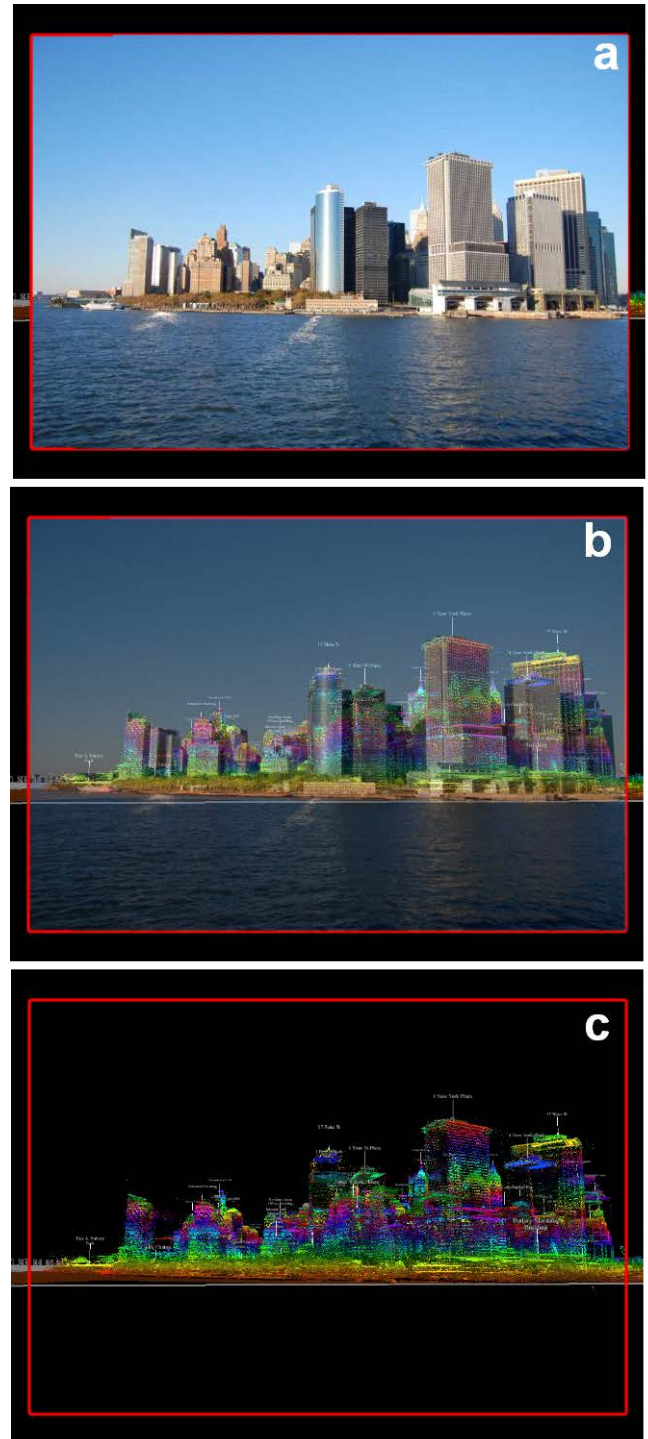

FIG. 9 ALIGNMENT BETWEEN A SKYLINE PHOTO AND 3D NYC MAP WITH (A) 0%, (B) 50% AND (C) 100% ALPHA BLENDING
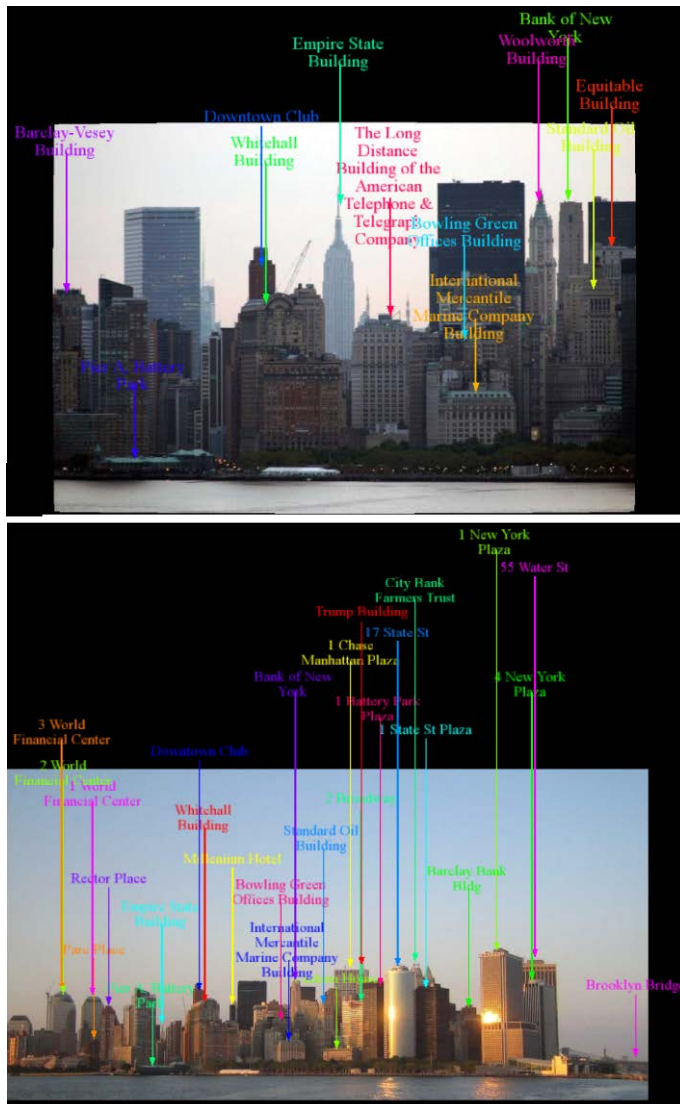
FIG. 10 SKYLINE PHOTOS AUTOMATICALLY ANNOTATED BY PROJECTING BUILDING NAMES FROM THE 3D NYC MAP
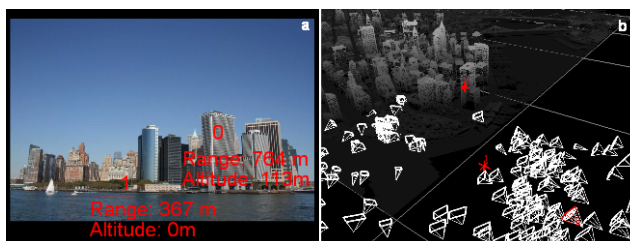

FIG. 11 IMAGE-BASED QUERYING. (A) USER SELECTS 2 FEATURES IN A GEOREGISTERED PHOTO. MACHINE SUBSEQUENTLY RAYTRACES THESE FEATURES BACK INTO THE 3D MAP. (B) POINTS CORRESPONDING TO SELECTED PHOTO PIXELS. POINT RANGES AND ALTITUDES ABOVE SEA-LEVEL ARE DISPLAYED INSIDE THE PHOTO WINDOW

in the NYC skyline photo on the left side of Figure 13 as belonging to sky, ocean or land. Once the 2D photo is georegistered, we can backproject each of its pixels into 3D space. If a raytraced pixel does not intersect any point in the 3D map (with occluding walls taken into account), we assume it corresponds to sky. Such identified sky pixels are tinted red in Figure 13b.

Pixels which backproject onto points with zero altitude above sea level are classified as ocean and are tinted blue in Figure 13b. Finally, all pixels exclusive from classification as sky or ocean are deemed to belong to land.   The result is a quite accurate segmentation of the image that is extremely simple to compute. While this particular algorithm may not work in all cases (e.g., places where water is above sea level), it can be easily extended to handle more detailed GIS data.

Image Retrieval

Finally, we demonstrate a simple image retrieval application. In particular, our system allows a user to enter the name of a building or landmark as a text string and receive a list of photos containing that object, ordered by some reasonable visibility criterion. We present here a simple version of such a gazetteer capability based upon projective geometry.

Exploiting the GIS layer within our 3D map for NYC, our machine first looks up the geolocation for a user-specified GIS label. After performing 2D fill and symmetry decomposition operations, it then fits a 3D bounding box around the ground target of interest. The computer subsequently projects the box into each georegistered image. In some cases, the bounding box
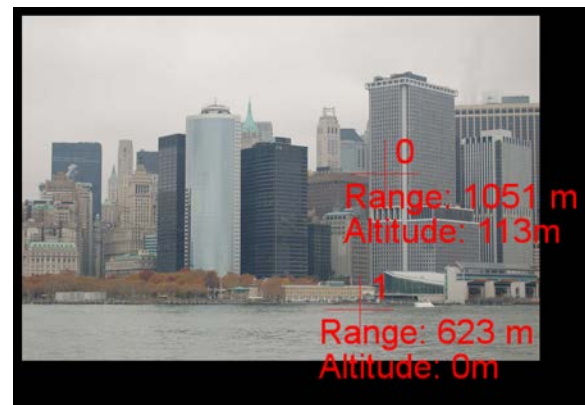

FIG. 12 POINTS IN FIG 11B REPROJECTED ONTO PIXELS IN ANOTHER GEOREGISTERED PHOTO. CAMERA RANGES TO FEATURES DEPEND UPON IMAGE, WHILE URBAN FEATURE ALTITUDES REMAIN INVARIANT

FIG. 13 PHOTO SEGMENTATION EXAMPLE. (A) GEOREGISTERED 2D PHOTO (B) AUTOMATICALLY IDENTIFIED SKY [OCEAN] PIXELS ARE TINTED RED [BLUE]

does not intersect a reconstructed camera's field-of-view, or it may be completely occluded by other foreground objects. But for some subset of the georegistered photos, the projected bounding box does overlap with their pixel contents. The computer then ranks the image according to a score function comprised of four multiplicative terms.

The first factor in the score function penalizes images for which the urban target is occluded. The second factor penalizes images for which the target takes up a small fractional area of the photo. The third factor penalizes zoomed-in images for which only part of the target appears inside the photo. And the fourth factor weakly penalizes photos for which the target appears too far off to the side in either the horizontal or vertical directions. After drawing the projected bounding box within the input photos, our machine returns the annotated images with sort according to their scores.

FIG. 14 illustrates the 1st, 2nd, 4th and 8th best matches to "Empire State Building" among our 1000+ reconstructed NYC photos. The computer scored relatively zoomed-in, centered and unobstructed shots of the requested skyscraper as optimal. As we would intuitively expect, views of the building for photos located further down the sorted list become progressively more distant and cluttered. Eventually, the requested target disappears from sight altogether. We note that these are not the best possible photos one could take of the Empire State Building, as our image database still covers a fairly small range of Manhattan viewpoints.
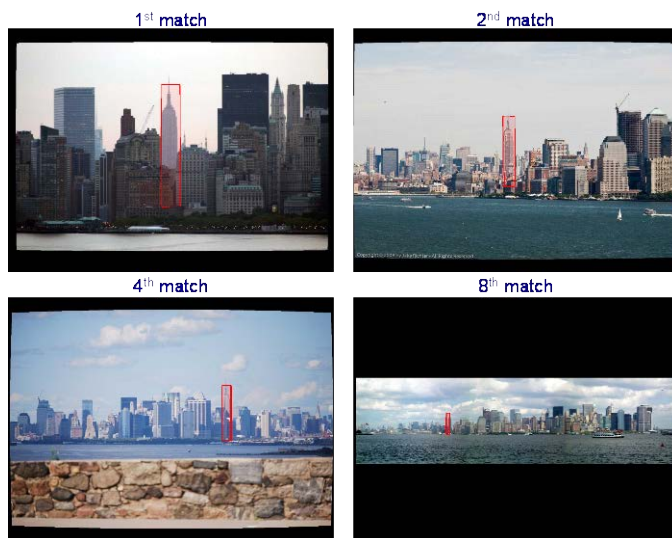

FIG. 14: FIRST, SECOND, FOURTH AND EIGHTH BEST MATCHES AMONG 1012 GEOREGISTERED NYC PHOTOS TO "EMPIRE STATE BUILDING". THE PROJECTION OF THE 3D BOUNDING BOX WITHIN EACH IMAGE PLANE IS COLORED RED

A similar retrieval capability was demonstrated in [1]. But in that work, objects were specified by selecting pixel regions within an image. Our system allows users to fetch photos via text queries instead.

Future robust versions of this image retrieval capability would provide a powerful new tool for mining urban imagery. Unlike current text-based search engines provided by Google, Flickr and other web archives, our approach requires no prior human annotation of photos in order to extract static objects of interest from complex city scenes. To the extent that input photos can be automatically reconstructed, the geometrical search technique is also independent of illumination conditions and temporal variations. It consequently can take advantage of the inherent 3D organization of all 2D photos for imagery exploitation.

Summary and Future Work

In this paper, we have demonstrated a prototype capability to reconstruct, georegister and enhance large numbers of uncooperatively collected urban digital photos. 3D photo reconstruction yields structured output from unstructured input. Ladar map registration augments photos with precise absolute geocoordinates and orientation metadata. And geometrical organization enables intuitive navigating and searching of large imagery archives.

Looking into the future, we foresee real-time interpretation of pixel outputs from mobile cameras operating in urban environments. As one walks, drives or flies through busy cities, information associated with buildings and streets could be used to automatically enhance an instantaneous image. Indeed, it should be possible to perform image-based queries of urban knowledge databases using photograph regions rather than text strings as inputs if accurate camera calibration can be rapidly calculated [26].

We are currently working with over 30,000 ground photos shot around the MIT campus in order to develop algorithms for mobile exploitation of urban imagery. Given its order-of-magnitude increase in size as well as its containing significant urban canyon occlusions, this next data set poses major new challenges beyond those discussed here for our initial 1000+ NYC photo experiments. But if technically feasible, near real-time annotation of live photo inputs would endow smart phones with powerful augmented urban reality capabilities. Given the dramatic rise in the quantity of digital images along

with camera phones over the past few years, the impact of systematically leveraging vast numbers of photos could someday rival that of Google text search.

REFERENCES

Arya, S. and Mount, D., "Approximate Nearest Neighbor Searching," in *SODA*, pp. 271-280 (1993).

Arya, S., Mount, D., Netanyahu, N., Silverman, R., and Wu, A., "An optimal algorithm for approximate nearest neighbor searching fixed dimensions," *Journal of the ACM (JACM),* vol. 45, pp. 891-923, (1998).

Bliss, N., Bond, R., Kepner, J., Kim, H., and Reuther, A., "Interactive Grid Computing at Lincoln Laboratory," *Lincoln Laboratory Journal,* vol. 16, pp. 165-216, (2006).

Bundler, http://phototour.cs.washington.edu/bundler/

Cho, P., "3D organization of 2D urban imagery," in *AIPR workshop*, pp. 3-8 (2007).

Cho, P., Bae, S., and Durand, F., "Image-based querying of urban knowledge databases," in *SPIE*, p. 733614 (2009).

Cornelis, N., Leibe, B., Cornelis, K., and Van Gool, L., "3D urban scene modeling integrating recognition and reconstruction," *International Journal of Computer Vision,* vol. 78, pp. 121-141, (2008).

Datta, R., Joshi, D., Li, J., and Wang, J., "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys,* vol. 40, (2008).

Ding, M., Lyngbaek, K., and Zakhor, A., "Automatic registration of aerial imagery with untextured 3D LiDAR models," in *CVPR*, pp. 1-8 (2008).

Fischler, M. and Bolles, R., "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM,* vol. 24, pp. 381-395, (1981).

Frueh, C. and Zakhor, A., "Constructing 3d city models by merging ground-based and airborne views," in *CVPR*, pp. II-562-9 (2003).

Grabler, F., Agrawala, M., Sumner, R., and Pauly, M., "Automatic generation of tourist maps," *SIGGRAPH Conf. Proc.,* vol. 27, p. 100, (2008).

Hu, J., You, S., and Neumann, U., "Integrating lidar, aerial image and ground images for complete urban building modeling," in *3DPVT*, pp. 184-191 (2006).

Kaminsky, R., Snavely, N., Seitz, S., and Szeliski, R., "Alignment of 3D Point Clouds to Overhead Images," in *CVPR*, pp. 63-70 (2009).

Kopf, J., Neubert, B., Chen, B., Cohen, M., Cohen-Or, D., Deussen, O., Uyttendaele, M., and Lischinski, D., "Deep photo: Model-based photograph enhancement and viewing," *SIGGRAPH Asia Conf. Proc.,* vol. 27, (2008).

Lowe, D., "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,* vol. 60, pp. 91-110, (2004).

Pollefeys, M., Nistér, D., Frahm, J., Akbarzadeh, A., Mordohai, P., Clipp, B., Engels, C., Gallup, D., Kim, S., and Merrell, P., "Detailed real-time urban 3d reconstruction from video," *International Journal of Computer Vision,* vol. 78, pp. 143-167, (2008).

Robertson, D. and Cipolla, R., "Building architectural models from many views using map constraints," in *ECCV*, pp. 155-169 (2002).

Russell, B., Torralba, A., Murphy, K., and Freeman, W., "LabelMe: a database and web-based tool for image annotation," *International Journal of Computer Vision,* vol. 77, pp. 157-173, (2008).

Snavely, N., Seitz, S., and Szeliski, R., "Photo tourism: exploring photo collections in 3D," in *SIGGRAPH Conf. Proc.*, pp. 835-846 (2006).

Sherrill, D., Yee, M., and Cho, P., "Analyzing networks of static and dynamic geospatial entities for urban situational awareness," in *SPIE*, p. 734605 (2009).

Snavely, N., Seitz, S., and Szeliski, R., "Skeletal graphs for efficient structure from motion," in *CVPR*, pp. 1-8 (2008).

Takacs, G., Chandrasekhar, V., Gelfand, N., Xiong, Y., Chen, W., Bismpigiannis, T., Grzeszczuk, R., Pulli, K., and Girod, B., "Outdoors augmented reality on mobile phone using loxel-based visual feature organization," in *MIR*, pp. 427-434 (2008).

Xiao, J., Fang, T., Tan, P., Zhao, P., Ofek, E., and Quan, L., "Image-based facade modeling," in *SIGGRAPH Asia Conf. Proc.* (2008).

Zebedin, L., Bauer, J., Karner, K., and Bischof, H., "Fusion of feature-and area-based information for urban buildings modeling from aerial imagery," in *ECCV*, pp. 873–886 (2008).

Zhao, W., Nister, D., and Hsu, S., "Alignment of continuous video onto 3D point clouds," *IEEE transactions on Pattern Analysis and Machine Intelligence,* vol. 27, pp. 1305-1318, (2005).

Peter Cho received a BS degree in physics from Caltech in 1987 and a Ph.D. in theoretical particle physics from Harvard in 1992. He worked as a postdoctoral physics research fellow at Caltech and Harvard from 1992 - 1998. Since joining Lincoln Laboratory in 1998, Cho's research has spanned a broad range of subjects in machine intelligence, computer vision and multisensor imagery fusion.

Noah Snavely is an assistant professor of Computer Science at Cornell University, where he has been on the faculty since 2009. He received a B.S. in Computer Science and Mathematics from the University of Arizona in 2003, and a Ph.D. in Computer Science and Engineering from the University of Washington in 2008. Noah works in computer graphics and computer vision, with a particular interest in using vast amounts of imagery from the Internet to reconstruct and visualize our world in 3D, and in creating new tools for enabling people to capture and share their environments. His thesis work was the basis for Microsoft's Photosynth, a tool for building 3D visualizations from photo collections that has been used by many thousands of people. Noah is the recipient of a Microsoft New Faculty Fellowship and an NSF CAREER Award.